Customised Profiling Model Development of Drivers to Predict the Risk of Involvement in a City Traffic Accident

Develop a predictive model to assess driver risk of traffic accidents using advanced machine learning techniques.

Introduction

1. Driver behaviors are pivotal in shaping traffic dynamics and safety, with risky driving behaviors significantly contributing to traffic accidents worldwide.
2. Recent studies have concentrated on building predictive models and datasets to comprehend and forecast risky driving behaviors.
3. Machine learning techniques like Support Vector Machines (SVM) are integrated into these models to create customized assessments of accident risks based on driver behaviors.
4. Rich datasets gathered from real-world driving scenarios are leveraged to train these models, providing a comprehensive understanding of various driving patterns and their associated risks.
5. The ultimate goal is to improve traffic safety by accurately predicting and mitigating potential accidents through proactive interventions based on driver behavior analysis.

Motivation

- Custom profiling model predicts drivers' accident risk, enhancing road safety by identifying potential dangers preemptively.
- Motivated by the urgent need to reduce global traffic accidents' toll, saving lives and mitigating economic and social costs.
- Integrating advanced machine learning and real-world driving data, our project aims to revolutionize road safety approaches.
- Empowering authorities, policymakers, and drivers with accurate risk assessments for safer and more efficient city transportation systems worldwide.

Aim

To develop a customized profiling model using machine learning techniques, specifically the Support Vector Machine (SVM) algorithm, to accurately predict the risk of drivers being involved in city traffic accidents based on their driving behavior data.

Problem Statement

Traffic Accident Statistics

21,510,650 traffic accidents caused by vehicles in 2016 in Taiwan.

Driver Fault

Significant accidents due to driver behaviours, both considerate and inconsiderate.

Challenges

Predicting accident risk is complex due to factors like road, weather, and natural disasters.

Data Description

Dataset

train_motion_data.csv, test_motion_data.csv

Features

Acceleration, gyroscope, class, timestamp

Data Sources

Accident reports, driver demographics, vehicle data, traffic and weather

test_motion_data

AccX	AccY	AccZ	GyroX	GyroY	GyroZ	Class	Timestamp
0.75819385	-0.2177906	0.45726347	0.0	0.0	0.0	AGGRESSIVE	818922
0.6675599	-0.03861046	0.23141575	-0.054367006	-0.0077121737	0.22525656	AGGRESSIVE	818923

SVM Algorithm Overview

1

What is SVM?

A supervised machine learning algorithm for classification and regression tasks

2

Hyperplane

A decision boundary that separates different classes in the feature space

3

Support Vectors

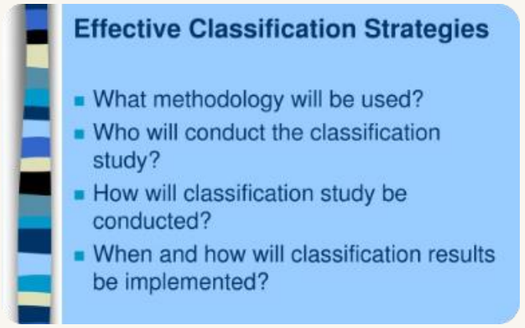
Data points closest to the hyperplane that influence its position and orientation

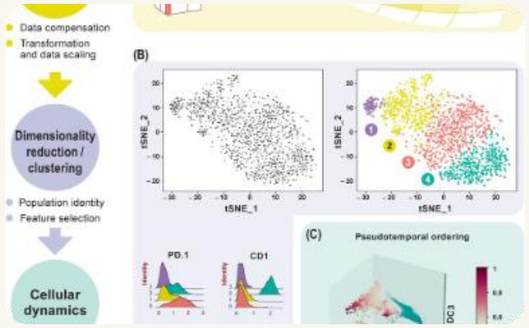
4

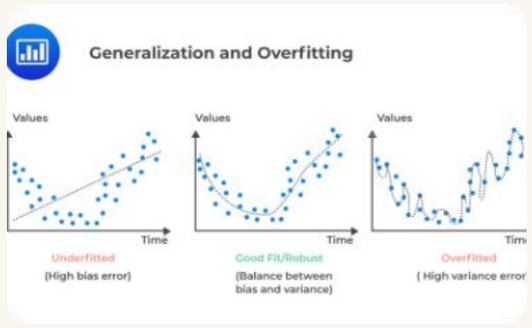
Margin

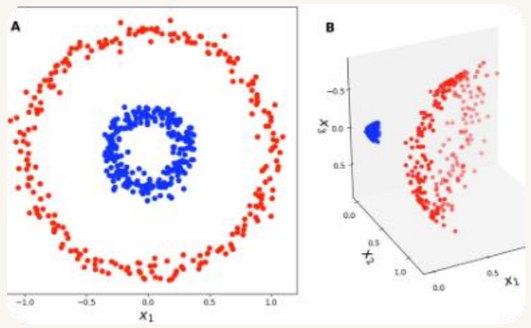
The distance between the hyperplane and nearest support vectors from either class

Why SVM









Effective Classification

SVM is known for high accuracy in binary classification tasks.

Handling High-Dimensional Data

SVM can manage complex datasets with multiple features.

Robust to Overfitting

SVM is less prone to overfitting with proper tuning.

Kernel Trick

Allows SVM to perform well with non-linear data.

How SVM is Used

1

Data Collection and Preprocessing

Gather diverse data, clean and preprocess to handle missing values and standardize features.

2

Feature Engineering

Create new features, normalize and standardize data for improved model performance.

3

Model Training

Split data into training and testing sets, apply SVM with appropriate kernels.

4

Model Evaluation

Evaluate performance using metrics, fine-tune SVM parameters to optimize the model.

5

Prediction

Use trained SVM model to predict accident risk for new drivers.

Model Training

1. **Data Preparation**
2. **Feature Engineering**
3. **Model Initialization**
4. **Training the Model**

The training process involves data preparation, feature engineering, model initialization, and training the SVM model. The linear kernel was chosen for its simplicity and effectiveness in high-dimensional spaces. The regularization parameter C was set to 1 to balance training and testing error.

Results Visualization

The accuracy score indicates the percentage of correctly predicted outcomes by our SVM model. With an our model demonstrates high performance in classifying drivers based on their risk of involvement in traffic accidents.

Accuracy	77.41%

Conclusion

Objective Achieved

Developed customized profiling model to predict traffic accident risk.

Effective Approach

Utilized SVM algorithm with linear kernel for high accuracy.

Robust Framework

Laid foundation for future research and practical implementations.

Key Findings

Comprehensive data crucial for reliable predictive models.